DOCUMENT RESUME

ED 151 078 PS 009 779

AUTHOR Hallau, Margaret Gardner
TITLE Improving Instructional Decisions in the Preschool Classroom.
PUB DATE 77
NOTE 89p.

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS Comparative Analysis; Criterion Referenced Tests; Decision Making; Educational Objectives; Instructional Improvement; Norm Referenced Tests; Observation; *Preschool Education; *Preschool Tests; *Testing; Test Reliability; *Test Reviews; *Test Selection; Test Validity

ABSTRACT

This paper discusses the use of assessment instruments at the preschool level and systematically reviews several measurement instruments used with young children. The first section briefly discusses the selection of educational objectives, consideration of alternative methods of collecting information about children, the selection of a specific instrument, and coordination among professionals. The second section, which contains the bulk of the paper, presents detailed analyses of 18 measurement instruments. Aspects of each instrument included in the analyses are: (1) publication information, (2) description, (3) validity, (4) examinee appropriateness, (5) reliability, replicability, and refinement, (6) teaching feedback, (7) useability, (8) retest potential and (9) summary. A series of charts comparing the instruments is also included. (JMB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED151074

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IMPROVING INSTRUCTIONAL DECISIONS IN THE PRESCHOOL CLASSROOM

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Margaret Gardner Hallau

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Margaret Gardner Hallau
SPED 459
Dr. Merle B. Karnes
Spring, 1977

PS 009279

IMPROVING INSTRUCTIONAL DECISIONS IN THE PRESCHOOL CLASSROOM

As the teacher in a preschool classroom plans for instructional activities, she or he asks many questions. "Can Tim plan a project and carry it out?" "Mary seems to have some difficulty cutting. Can most three-year-olds cut on a line?" "Sue can name objects. Can she describe them?" This paper will explore some of the ways preschool teachers answer questions about children and will describe some ways instructional decisions in the preschool classroom can be improved. In the first section of the paper, the focus of discussion will be the selection of objectives, consideration of alternative methods of collecting information about children, the selection of a specific instrument, and coordination among professionals. The second section includes detailed analyses of eighteen instruments often used to make decisions about young children. For easy reference and comparison among the instruments, a series of charts comparing the pertinent information from each instrument follows the analyses.

Rowan (1973) notes that the development of children is an evolving process, difficult to measure with standardized instruments. Even though the development of the child is difficult to measure, others (Snelbecker, 1974 and Thorndike, 1967) are concerned with the unsystematic way in which educational decisions are made. Much of what is done seems to stem from common sense rather than an overall conception of educational objectives and the best means for attaining them. Thorndike (1967) suggests that decisions imply prediction; measurement helps teachers make predictions with smaller margins of error. Since predictions become more fallible with an increase in time between predictions based on the measurement of a behavior and the

actual outcome, teachers should collect current evidence when important decisions are to be made, especially for young children.

Selection of Educational Objectives

The framework in which a teacher makes decisions should be determined by the objectives of the classroom. Only after objectives have been determined can effective procedures and instruments be selected assess how well the children are meeting the objectives. Gottman and Clasen (1972) note that educators often begin planning an educational program at the "how" level. If no objectives are determined, it is unlikely that an adequate search will be made for alternative procedures, including assessment procedures. Bloom (1971) suggests the teacher consider both what objectives are desirable and what objectives are possible. Objectives can be determined from a variety of sources. Snelbecker (1974) outlines a variety of educational and psychological theories which could form the basis for selecting objectives. Of direct relevance to the preschool educator is Kamii's (1971) discussion of objectives to help young children in socio-emotional, perceptual-motor, cognitive, and language development.

Those preschool programs serving handicapped children have additional contingencies on which instructional decisions must be based. Under P.L. 94-142, each handicapped child must have an individual educational program which specifies the child's current educational performance as well as the annual goals and short term instructional objectives. In addition, the parents and representatives from the school will review the plan. The teacher in a preschool which serves handicapped children, then, must be familiar with objectives appropriate for children with particular handicapping conditions as well as objectives appropriate for all children.

Under the Congressional mandate of 1972, ten percent of the slots in Head Start are reserved for children with handicaps. According to the Performance Standards against which all Head Start programs are evaluated, each child must have individual objectives designed to meet his or her unique needs. Since under P.L. 94-142, all handicapped children are the responsibility of the state education agencies, Head Start agencies in states with mandatory legislation for early childhood are accountable for individual objectives under two separate legislative mandates. If objectives are not determined prior to selecting an instrument to measure those objectives, it is possible that the assessment process, rather than providing appropriate educational activities, may become the focus of the educational program. The "perfect" instrument which will measure all the competencies needed by young children has not been developed, nor is it likely to be with children being as delightfully complex as they are. Rather than select an assessment procedure first, the teacher should first clarify the objectives children should attain and then select assessment procedures which will help the teacher determine how the children are progressing towards the desired objectives. While each child has individual objectives tailored to meet his or her specific needs, all objectives are from the domain of objectives which are appropriate for children's growth and development.

An assumption of this procedure is that teachers have not only a knowledge of child development, but also skills in planning and implementing individual programs for children. If teachers in a particular group, be it a Head Start agency or any other group which serves young children with handicaps, do not possess the competencies required to plan and implement programs, it is the responsibility of the administrators to provide training to the staff (or possibly locate new staff). No assessment procedure can

replace the observations and insights of a trained teacher who is familiar with the child's usual behavior.

Consideration of Alternative Methods of Collecting Information About Children

Once the teacher has determined which objectives will receive major focus, the next step is to determine where the children are in relation to the objectives. There are a variety of strategies available to obtain information about children. The teacher might use some of the observational approaches such as the case study approach suggested by Rowan (1973) or the event sampling suggested by Medinnus (1976). If the teacher is interested in recording categories of behavior, the 73 observation systems reviewed by Boyer, Simon, and Karafin (1973) should be explored. Evans (1974) notes that currently there is less of an emphasis on using formal tests alone and more of a trend toward systematic observation.

Another strategy for obtaining information on which to base instructional decisions is to administer a test. Cronbach and Gleser (1967) suggest that educational tests exist because our society confronts people with decisions for which they have insufficient information. They believe the value of a test depends on the accuracy of the instrument, its relevance to the decision being made, and the loss resulting from an erroneous decision. Ellis (1974) and others suggest that the most important question is "Will this particular technique measure what I want to measure?" Bloom (1971) states that standardized diagnostic tests can be used to identify weaknesses in prerequisite skills and abilities. Patterns of responses serve as the basis for interpretating diagnostic tests.

There are several difficulties associated with the use of standardized tests. Among the potential abuses discussed by Ebel (1964), the most harmful

is that subsequent academic or social status may be determined by the way expectations become established among those who have access to the test results. The effects labelling children have been noted, particularly in recent years among special educators. Anastasi (1976) points out that systematic variations among scores obtained by the same individual on different tests may be due to factors, such as incomparable scale units, which are not related to changes in the individual. Anastasi also outlines the problems in administering tests to preschoolers: normative samples are usually smaller and less representative than for school age kids; reliability is usually lower than for older children. In addition, an examiner who is alert to problems of testing young children is usually required. Another difficulty with standardized tests is that some assume a statistical model in which scores are distributed in a bell-shaped curve. Meier (1973) suggests that the multi-culturally aspects of our society are not represented by the bell-shaped curves.

With due consideration for the difficulties noted above, assessment instruments can provide information needed to make appropriate decisions about children. The potential test or assessment instrument user might want to read some evaluations of instruments prior to selecting instruments for classroom use. Johnson and Bommarito (1971) have developed a handbook of unpublished instruments which appeared in journals between 1956 and 1965. The measures described in their publication were selected because they were suitable for use with children up to age 12; they were available to other professionals; they had enough scoring and administration information to enable others to use them; and the measures were long enough that norms and meaningful reliability could be developed. Another excellent source of information about tests is the CSE/ECRC Preschool and Kindergarten Test Evaluations (1972) developed by the Center for the Study of Evaluation at

UCLA. Over 125 tests are categorized by concepts which the tests measure and then each is evaluated systematically. Buros' Mental Measurements Yearbook contains test evaluations; however, only a quarter of the tests reviewed for this paper were reviewed in Buros. Other sources of information about tests are the Educational Resources Information Center (ERIC) and journals such as the Journal of Educational Research and Educational and Psychological Measurement. Groups such as TADS may publish bibliographies of tests; however, evaluations are seldom included in such documents.

Selection of a Specific Instrument

There are many ways to classify tests and other assessment procedures. One of the most useful and necessary distinctions is to consider whether an instrument is criterion-referenced or norm-referenced. According to Anastasi (1976) criterion-referenced instruments attempt to give content meaning to scores by interpretating the individual's performance in terms of what he can do. Often the developers of criterion-referenced tests point out that traditional testing issues such as validity and reliability are not as important in the evaluation of criterion-referenced tests because the performance of the individual is not being compared to the performance of other individuals but rather is being compared to a standard. The potential test user can examine the items and determine if the items would assess what he or she wishes to assess. A second major classification of instruments include those which are norm-referenced. The scores of individuals are ranged along a continuum based on their relative location in the distribution of others who have been measured on the characteristic being studied (Anastasi, 1976). Issues are validity and reliability are central to understanding the use of norm-referenced instruments.

Thorndike (1971) notes that the test itself is not validated. Rather, it is the interpretation of data which arises from a certain procedure which is verified. Thus, as is also noted by the American Psychological Association, validity information indicates the degree to which a test is capable of achieving certain aims. Various validity coefficients relate to different aims of testing. According to the standards for tests and manuals that were compiled by the APA, content validity is demonstrated by showing how well the test content samples behaviors about which conclusions are to be drawn. Criterion-related validity may be demonstrated by comparing the test scores with external variables considered to provide direct measure of the behavior in question. Construct validity describes what the test measures by determining the degree to which certain explanatory constructs account for performance on a test. Often construct validity is measured by correlating one test with another test. Harrocks (1964) points out that the problem of establishing the validity and reliability of the criterion may be as great as establishing them for a particular measuring instrument.

The reliability of an instrument is another important consideration. In the standards for tests and manuals developed by the American Psychological Association, reliability refers to the consistency and stability of measurement by a test. Several methods may be used to derive the instrument's reliability. The scores on one set of measures might be correlated to scores on an alternate form. Or, the odd and even items on a test with twice the number of items can be correlated. Part of the study of reliability is the study of factors, such as variations in administration, which contribute to inconsistency in observations. The teacher who is selecting an instrument for use in the decision-making process, must consider carefully the reliability and validity of the tests under consideration. As Tyler (1971) points out, if a test does not measure what it purports to

measure, it is worse than useless because the user will act on the wrong assumption about what a score means.

It is also important that the sample on which the norms of the test, if the instrument is a norm-referenced measure, be large enough to provide stable values and be representative of the population under consideration (Anastasi, 1976). Any reported scores should be accompanied by a specification of the test from which the scores were derived.

Other factors should be considered in selecting a test. For example, how much time is needed to administer the test? Is it to be administered individually or to a group? How much training must the examiner have? What materials are needed? How much does the test cost? This information, along with adequate information to permit evaluation of the test itself, should be included in the manual (Anastasi, 1976).

In addition, the method of recording scores and then interpretating the scores should be considered. Raw scores are meaningless without information on which the scores can be interpreted. In many cases, a psychologist or other measurement specialist will be needed to interpret a particular test score. Many instruments, particularly those selected for review in this paper, can be used by teachers to make instructional decisions. The teacher, however, must be familiar with the instrument and be able to utilize the information in planning for children. The information from test scores might be thought of as alternative hypotheses about why a particular child exhibits a particular behavior. The teacher must continue to gather evidence until a decision can be made. Then the teacher must continue to moniter the outcomes of a particular instructional decision and alter her instruction if necessary.

Coordination Among Professionals

In order for the child to receive the benefits of the determination, description, and classification of his or her behavior that occurs as a part of the screening, assessment, diagnosis, treatment process, it is important for all those who have information about the child to coordinate their efforts. In 1970, Fiapan, Gunn, and Neubauer compared the development assessments of 45 young children. The sources of information included mother interviews, teacher interviews, child interviews, observation of the child in school, and psychological testing. They concluded that school observations provided the most favorable assessment for boys, child interviews the most favorable for girls, and psychological testing for both boys and girls.

Particularly in the case of handicapped children, the teacher and auxillary personnel such as psychologists and speech therapist collect information about the child. Each can learn from the other if they have the opportunity to share their information about the child. In some cases, psychologists may not be aware of innovations in curriculum which would be appropriate for a particular child. The teacher, on the other hand, may be unfamiliar with some of the terms used by the speech therapist or be unfamiliar with certain kinds of scores. Ebel (1967) suggested that in order to be competent in educational measurement the teacher should know, among other things, the educational uses and limitations of tests, should know the criteria by which the quality of a test can be judged, should know how to select and administer a test, and should know how to interpret test scores correctly, but with recognition of their limitations.

If the efforts of the professionals are not coordinated, it is possible that the child may be assessed on a skill several times. Part of that time could be more effectively spent in learning something new.

The lack of coordination of professionals is sometimes a problem in Head Start agencies when they are providing services to handicapped children. The Head Start teacher may be unfamiliar with some of the terms on the reports which follow the child back from his "diagnostic evaluation." The professional who provided the evaluation may be unable to make other than general suggestions about what to do with the child once he or she is back in the classroom. On the other hand, some Head Start agencies have developed procedures which insure that staffings occur which attended by all those who make decisions about the child. One strategy found useful by some of the more successful Head Start agencies is to select an instrument which meets the objectives of the program and then to train the staff to collect systematic information about the children.

In summary, the collection of information necessary to make instructional decisions about children is a necessary competency for pre-school educators. The discussion in this section has focused the selection of assessments in terms of the objectives of the classroom (or program) and in terms of the characteristics of the instrument itself.

Analyses of Instruments

In order to evaluate the instruments in a consistent, systematic manner, a test evaluation process was developed after a review of the literature. The process is an extension of the MEAN approach used by the Center for Evaluation in its review of preschool and kindergarten instruments. It was felt that specification of information in each category rather than a numerical rating would be important in making a decision about the use of a particular instrument. Therefore, a series of questions, modified from criteria utilized by the Center for Evaluation to evaluate cognitive, affective, and interpersonal skills, were generated. The questions include information on all important aspects of a test. Where appropriate, criteria are also indicated in parentheses after the question.

The tests analyzed in this section were selected on the basis of the following criteria:

(1) cited in at least two bibliographies of tests for young children, or

(2) listed by International Training Consultants as probably being used by Head Start centers. An * indicates those tests selected on the basis of the second criteria.

The instruments reviewed include the following:

ABC Inventory

*Assessment by Behavior Rating (ABR)

Basic Concept Inventory

Carolina Developmental Profile

Circle Preschool Individual Assessment

CIP

*CIRCUS

*Cooperative Developmental Scale (Project MEMPHIS)

*Denver Developmental Screening Test

*Developmental Progress Scale

*Developmental Screening Questionnaire

*DIAL

*Koontz

*LAP

*Peabody Picture Vocabulary Test

*Preschool Attainment Record (PAR)

*Santa Clara Inventory

After the individual analyses are a series of charts which provide comparative data on a number of critical variables.

Title: ABC Inventory to Determine Kindergarten and School Readiness (Research Edition)

Authors: Norman Adair and George Blesch

Publisher: Research Concepts
1368 East Airport Road
Muskegon, Michigan 49444

Description:

The ABC Inventory is an individual school readiness test, orally administered in four sections: simple drawing of a man, naming, common knowledge, and counting skills. The recording form contains suggested answers and indicates the number of points to be awarded. This instrument provides a brief inventory of four categories of behavior related to school readiness. The age range is 3 years 6 months to 6 years 7 months. The cost is $4.00 per 50 record forms; $1.00 for a specimen set. (1965)

VALIDITY

What does this test measure?
This test measures readiness for school.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)
The ABC Inventory contains a total of 18 items. With so few items, it is important that each one be maximally discriminative. Several items, including the number of wheels on a motor scooter and "What is ice when it melts?" are confusing and irrelevant.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)
The ABC Inventory seems to identify more often children who fail in first grade as compared to children who pass first grade. The authors mention that item analyses were conducted between 1960 and 1962 but no details are given.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?
Yes.

Are instructions to the child about what he is to do and why clear and suitable in tone?
Yes.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)
Yes.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?
The recording from provides adequate space for the child's drawing.

Are the length of the test and the time pressures appropriate?
The Inventory is very brief; probably too brief to be used for more than a screening instrument.

How are answers recorded?
The answers are recorded by the examiner directly on the recording form.

What equipment is needed?
Paper squares, pencil.

Is it readily available?
Yes.

How much maintenance does it require?
None.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)
Reliability was established by matching comparable groups and assuming group equivalency. No indication is given of the variables on which the groups were matched. Even though the means of the test scores are not significantly different, the groups cannot be assumed to be equal in light of the lack of specification of the variables on which the groups were matched.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)
No alternate forms.

Are procedures for replicating statistical procedures clearly described?
No indications are given of the criteria on which the groups were matched. It would be impossible to replicate the reliability study of this instrument.

How are converted scores reported?
The raw scores are converted to Ready Ages, an approximate of the Stanford-Binet M.A.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)
The norm range is distressingly limited: 4 years 9 months to 4 years 11 months. Scores are not separated by sex or socio-economic status.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?
The norms were developed in 1964. Large school districts are not represented. No indication is given of the geographic or ethnic makeup of the norming group.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?
Yes.

Is special training required to interpret the test scores?
No.

Can this test be directly, clearly, immediately interpreted?
Yes.

Does the test provide guidelines for making decisions from a score or score interpretation?
The manual provides a chart for interpreting the scores of children under five in terms of the child's probable success in kindergarten. However, since the ABC Inventory identified correctly only 63% of the students who passed their first year of school, the chart is of limited usefulness.

USEABILITY

Is this test administered to individuals or groups?
Individuals.

How much time does administration take?
Usually eight to nine minutes.

Does the manual contain a discussion of the purpose, use and limitation of the test?
Yes, though the limitations are not explored to a significant degree.

Does the manual contain clear administration and scoring directions?
Very explicit directions are provided.

Does the manual provide a description of test development, analysis procedures, and normative groups?
The description is too brief to allow either replication or analyses.

Can the test be administered by the teaching staff without extensive practice?
Yes.

How much time is necessary to prepare for the test administration?
Under 5 minutes.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

The ABC Inventory is of limited usefulness to the preschool educator. Its brevity and ease of administration cannot compensate for the lack of an adequate number of items, items of questionable validity, and serious statistical errors in developing measures of reliability. As rated in the CSE-ECRC Preschool/Kindergarten Test Evaluations, it received a "poor" for "measurement validity" and "normed technical excellence" and a "fair" in "examinee appropriateness" and "useability."

Title: Assessment By Behavior Rating

Author: Elizabeth Sharp

Publisher: Model Preschool for Handicapped Indian Children
Department of Special Education
University of Arizona
Tucson, Arizona 85721

Description:

Assessment By Behavior Rating is an instrument to help the children differentiate between a child's developmental strengths and weaknesses in physical development, self-help skills, language development, social and emotional development. In each area are descriptors of behaviors associated with the area. Each description is divided into 5 levels arranged from simple to complex. For example, under "physical development," the use of scissors ranges from "unable to use scissors" to "uses scissors well and enjoys cutting." Three sheets to aid in interpretation are attached to the back of the instrument. No indication is given as to the age of the child for which this instrument is appropriate except that the levels range from age one to age five. (1974)

VALIDITY

What does this test measure?

This instrument measures physical development, self-help skills, language development, social and emotional development but only in a very general sense.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

The five items in physical development assess only five motor skills such as cutting. The care-of-self section omits toileting. Language development contains one item (with five levels) for each category associated with the ITPA. Under social and emotional development, seventeen behaviors are listed. Each is delineated on a five point scale from "never" to "always." The behaviors, listed in general terms, may be interpreted differently by different examiners.

What empirical evidence exists that would give an administrator confidence in its usage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

None is listed.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

Yes. In addition, many items can be observed during regular classroom activities.

Are instructions to the child about what he is to do and why clear and suitable in tone?

For some items, a general suggestion is made about instructions to the child.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)
The printing is clear.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?
In its present form, the instrument is not produced commercially.

Are the length of the test and the time pressures appropriate?
There is no indication of how long the instrument takes to administer or the conditions under which it is to be administered.

How are answers recorded?
A check mark is placed under the best description of the functioning level of the child.

What equipment is needed?
Pencil, test materials are not specified but are suggested by the behavior described in the categories.

Is it really available?
Unknown.

How much maintenance does it require?
Unknown.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)
Not stated.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)
Not stated.

Are procedures for replicating statistical procedures clearly described?
No indication is given of any reliability studies. This is an extremely serious omission in light of statements on the Interpretation Sheet that indicate a tentative diagnosis may be made in the area of retardation if the child is functioning two years below his chronological age in all areas of growth and development. The Interpretation Sheet also indicates that a diagnosis in the area of learning disabilities may be made if there are discrepancies of at least one year between functional levels in growth and development.

How are converted scores reported?
Scores are converted into a language profile and a developmental age.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)
None indicated.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?
None indicated.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?

The developmental age for a particular area is determined from the largest total of checks per column. A language profile is constructed from the ratings in the language area.

Is special training required to interpret the test scores?

No mention is made of the training required; however, the lack of objectivity of the items would seem to require a trained examiner.

Can this test be directly, clearly, immediately interpreted?

The "Interpretation Sheet" seems straightforward, but not enough items are included in each category for adequate interpretation.

Does the test provide guidelines for making decisions from a score or score interpretation?

Yes, but the tentative diagnosis of children on the basis of this instrument is unwarranted without any information on the validity or reliability of the instrument.

USEABILITY

Is this test administered to individuals or groups?

Individuals.

How much time does administration take?

Unspecified

Does the manual contain a discussion of the purpose, use and limitation of the test?

No.

Does the manual contain clear administration and scoring directions?

Directions occasionally accompany individual items. No general directions for administration are given. Scoring directions are straightforward.

Does the manual provide a description of test development, analysis procedures, and normative groups?

No.

Can the test be administered by the teaching staff without extensive practice?

Training in observation seems to be necessary.

How much time is necessary to prepare for the test administration?

Unable to determine since inadequate directions for administration are specified.

RETEST POTENTIAL

Are alternate forms of the test available?

Unknown.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?

Unknown.

SUMMARY

Assessment By Behavior Rating is an instrument of limited usefulness in the hands of an untrained examiner. It should be used cautiously, if at all, by a trained examiner. The behaviors listed lack the degree of specification necessary for valid decisions to be made about children, particularly since no validity or reliability data are reported. Further, all the areas except Emotional Development have an inadequate number of items with which to assess the child's level of functioning.

Title: Basic Concept Inventory

Author: S. E. Engleman

Publisher: Follett Educational Corporation
Chicago, Illinois

Description:

The Basic Concept Inventory is an individually administered, criterion referenced checklist of basic concepts used in explanations and instructions in first grade. Designed for children beginning academic tasks, it includes three sections: basic concepts, statement repetition and pattern awareness. Picture cards are used in the first section. The other two sections require verbal responses. The manual contains a comprehensive description of the utilization of test results in programming for children. The test costs approximately $8.00 for a manual, picture cards, and 15 record forms. (1967)

VALIDITY

What does this test measure?

The Inventory attempts to measure a child's familiarity with conventional statements and a child's perception of auditory sequences.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

Basic concepts such as "not," "and," "or" are included. A listing of all concepts would facilitate the teacher's interpretation in the score. The manual provides extensive information on each item but lacks a brief summary of all concepts under study.

What empirical evidence exists that would give an administrator confidence in its usage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

Extensive description is provided for the use of criterion referenced tests. The author suggests that performance on the Inventory be treated as a work sample for making instructional decisions. The Buros' reviewer states that the goals are worthwhile and the inventory plan is potentially valuable.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

Yes, though the emphasis is on verbal items.

Are instructions to the child about what he is to do and why clear and suitable in tone?

The child is given clear instructions but no reasons.

Are the printing and/or illustration of the appropriate style, size, and clarity?

The pictures include individuals from several racial backgrounds.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?
Adequate space is provided.

Are the length of the test and the time pressures appropriate?
The test is untimed and is of suitable length.

How are the answers recorded?
The examiner records them on the answer sheet.

What equipment is needed?
Pencil.

Is it readily available?
Yes.

How much maintenance does it require?
None.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)
The author discusses the construction and use of norm-referenced tests in contrast to criterion-referenced tests. No reliability data is reported. However, the author notes that the ABC Inventory is a tool to be used by the teacher to get a sample of the child's work so that decisions regarding classroom instruction can be made.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)
No alternate forms.

Are procedures for replicating statistical procedures clearly described?
No statistical procedures are reported.

How are converted scores reported?
Converted scores are not computed.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)
The author states that age norms are not the focal importance of this criterion referenced instrument because norms do not indicate what a particular child has learned.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?
Norms are not indicated.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?
Converted scores are not computed.

Is special training required to interpret the test scores?
No. The author provides an extensive section on the interpretation of the scores, including many examples and case studies.

Can this test be directly, clearly, immediately interpreted?
A child's incorrect answer to a particular item could be related to several errors in the child's conceptual framework. The teacher would need to look at several items to determine the error pattern.

Does the test provide guidelines for making decisions from a score or score interpretation?
The suggestions for making instructional decisions about a child are one of the most helpful features of this Inventory. The second part of the manual provides a comprehensive introduction to instructional strategies associated with precision teaching.

USEABILITY

Is this test administered to individuals or groups?
Individuals.

How much time does administration take?
Approximately 20 minutes.

Does the manual contain a discussion of the purpose, use and limitation of the test?
An excellent discussion is provided.

Does the manual contain clear administration and scoring directions?
Yes.

Does the manual provide a description of test development, analysis procedures, and normative groups?
Information in these areas is lacking.

Can the test be administered by the teaching staff without extensive practice?
Yes.

How much time is necessary to prepare for the test administration?
A minimum of preparation time is needed.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

If the ABC Inventory is used as the author intended, that is, to provide work samples to the teacher to aid in instructional decision-making, then the Inventory can provide some information to the teacher about the child's level of mastery of concepts such as "and," "but," "not" and plurals. Other tests should be used for diagnostic purposes.

Title: Carolina Developmental Profile

Author: David L. Lillie and Gloria L. Harben

Publisher: TADS
625 W. Cameron Avenue
Chapel Hill, N.C. 27514

Description:

The Carolina Developmental Profile is a criterion-referenced checklist designed to help the teacher establish long-range objectives in these six areas: fine motor, gross motor, visual perception, reasoning, receptive language, and expressive language. Each task includes a task description, developmental age, needed materials, and criteria for passing. The age range is 2 years to 5 years. (no date indicated)

VALIDITY

What does this test measure?

The Carolina Developmental Profile provides information on motor, language, reasoning, and perception.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

The number of items under each area varies from twelve to twenty. The items seem to tap a variety of behaviors.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

The author states that the items were developed after a careful review of the literature and extensive testing on children. Further specification, unfortunately, is not included.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

A child would probably find the items interesting and fun.

Are instructions to the child about what he is to do and why clear and suitable in tone?

No instructions to the child are indicated.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)

The illustrations are simple yet appealing.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

Yes,

Are the length of the test and the time pressures appropriate?
The Profile is untimed and should be given in several sessions.

How are answers recorded?
The teacher and the child both record answers on the Profile.

What equipment is needed?
Yes, some commonly available classroom materials are needed for some items.

Is it readily available?
Yes.

How much maintenance does it require?
Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)
None is indicated.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)
None is indicated.

Are procedures for replicating statistical procedures clearly described?
None are indicated.

How are converted scores reported?
The raw scores are converted to a Developmental Age Ceiling based on the number of tasks he can do in a particular area.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)
None are indicated.

Are the norms current (within 5 years) and representative of geopgraphic areas, age, racial and ethnic origins, types of schools and school districts?
None are indicated.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?
Yes, only simple addition is required.

Is special training required to interpret the test scores?
No, the scores provide information for planning long range objectives. The teacher must have other information to plan a comprehensive, individual program since the items do not tap competencies in a particular area.

Can this test be directly, clearly, immediately interpreted?
Yes, with the limitation indicated above.

Does the test provide guidelines for making decisions from a score or score interpretation?
Yes, the items missed become long-range objectives for the child.

USEABILITY

Is this test administered to individuals or groups?
Individuals

How much time does administration take?
Not indicated.

Does the manual contain a discussion of the purpose, use and limitation of the test?

Yes, the information in these areas is specific.

Does the manual contain clear administration and scoring directions?
Yes, each task has a criterion level listed.

Does the manual provide a description of test development, analysis procedures, and normative groups?
A brief statement concerning the careful review of the literature and extensive testing on children are provided as is a list of 17 references. No description of analysis procedures or normative groups is provided.

Can the test be administered by the teaching staff without extensive practice?
Yes, if the staff has been trained to observe children.

How much time is necessary to prepare for the test administration?
Minimal.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

The Carolina Developmental Profile can help the teacher determine long range objectives for children. In order to be able to utilize the Profile effectively and not just "teach to the task," the teacher must be skilled in task analysis. This instrument is not intended to be used for a comprehensive diagnostic evaluation. The validity of the instrument is not indicated, nor is the reliability. While the author does not believe that this information is necessary for a criterion-referenced test, those who make decisions on the basis of this instrument would be better served if the information were available.

Title: Circle Preschool Individual Assessment

Author: Glenda Frankhauser, Alicia Fazio, Nancy Giessman, Jack Harley, Suzanne Hering, and Annie Oliver

Publisher: Circle Preschool
9 Lake Avenue
Piedmont, CA. 94611

Description:

The Circle Preschool Individual Assessment, a synthesis of other assessments used in early childhood programs, accompanies the Live Oak Curriculum. It provides information to the teacher about these six areas: language, cognitive, gross motor, fine motor, socio-emotional, and self-help. The specific skill areas in the language section reflect the areas of the ITPA model. The teacher observes the child to obtain the required information. The Individual Assessment items are indexed in the Curriculum Guide to specific lesson plans. The range of the Individual Assessment is 12 months to 72 months. The cost is $20.00 for the Curriculum Guide and accompanying Individual Assessment. (1976)

VALIDITY

What does this test measure?
This instrument measures the child's level of functioning in a variety of areas.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)
The Assessment provides a wealth of information on an individual child. Each age period has several items with developmental landmarks noted. No validity information is presented, unfortunately.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)
No information is presented as to how the items were selected or ordered.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?
Many of the items can be observed during regular classroom activities.

Are instructions to the child about what he is to do and why clear and suitable in tone?
No instructions are specified.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)
Yes.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

In its present form, the Assessment is 44 pages. It would be cumbersome to try to assess several children, at the same time, as the manual suggests, without taking notes and then transferring the information to each child's individual form.

Are the length of the test and the time pressures appropriate?

The Assessment is untimed and will be administered over a period of time.

How are answered recorded?

The teacher records the answer on the form.

What equipment is needed?

No equipment list is provided but a variety of items, usually found in a classroom, are needed for particular items.

Is it readily available?

Yes.

How much maintenance does it require?

Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more admministrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

No reliability measures are indicated.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)

No alternate forms are provided.

Are procedures for replicating statistical procedures clearly described?

No statistical procedures are indicated.

How are converted scores reported?

Scores can be transferred to Profiles.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)

None are reported.

Are the norms current (within 5 years) and representative of geographic areas, age, racial and ethnic origins, types of schools and school district?

None are reported.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?

Since not all tasks have criterion levels indicated, some tasks may be subject to several interpretations. No information about the number of items which must be passed for an age level to be considered passed on the profile are indicated.

Is special training required to interpret the test scores?
No. The scores are to be used by teachers in planning objectives for the child.

Can this test be directly, clearly, immediately interpreted?
Yes.

Does the test provide guidelines for making decisions from a score or score interpretation?
Yes. After the child makes three errors, the teacher stops assessment and begins instruction.

USEABILITY

Is this test administered to individuals or groups?
For some tasks, several individuals may be observed at once.

How much time does administration take?
The administration time is unspecified.

Does the manual contain a discussion of the purpose, use and limitation of the test?
Yes. The authors note that the Individual Assessment suggests an overall guideline for sequencing the content of classroom activities based on a developmental model.

Does the manual contain clear administration and scoring directions?
The administration procedures are general. No specific information is provided on the scoring of individual items.

Does the manual provide a description of test development, analysis procedures, and normative groups?
None is provided.

Can the test be administered by the teaching staff without extensive practice?
Yes.

How much time is necessary to prepare for the test administration?
Minimal.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

If used as the developers intended, the Circle Preschool Individual Assessment provides information the teacher can use in planning for the sequencing of content. Unfortunately, the items do not contain standardized scoring instructions and items may be scored differently, either at different items or by different people. The Profile constructed from these measures might not accurately reflect the child's accomplishments. The accompanying Curriculum Guide contains hundreds of activities to which the Individual Assessment has been indexed.

Title: CIRCUS

Authors: Scarvia B. Anderson, Gerry Ann Bogatz, Thomas W. Draper, Ann Jungeblut, Gary Sidwell, William C. Ward, Allen Yates

Publisher: Educational Testing Service
Princeton, New Jersey

Description:

The 17 instruments which comprise CIRCUS include measures of receptive vocabulary, quantitive concepts, visual discrimination, perceptual-motor coordination, letter and numeral recognition and discrimination, discrimination of real world sounds, auditory discrimination, aspects of functional language, comprehension, interpretation, and recall of oral language, productive language, general information, visual and associative memory, problem solving, divergent pictoral production, an activities inventory, behavior inventory, and educational environment questionnaire. Thirteen of the fourteen direct child measures can be administered to small groups. Users may select those instruments which best fit their assessment needs. Scores are reported in "sentence reports" which describe the child's behavior. The age range includes 4 and 5 year old children. (1974)

VALIDITY

What does this test measure?
Language, cognitive, perceptual motor skills.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)
Each instrument provides sufficient opportunity for the child to demonstrate his understanding of the task. Eleven of the measures, however, require the same response format, marking the correct picture.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)
This is a criterion-referenced measure with a sample of almost 3,000 children in the standardization sample. The Technical Manual describes content and other analyses.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?
Children would probably enjoy the booklets.

Are instructions to the child about what he is to do and why clear and suitable in tone?
The suggested instructions provide excellent guidance to the teacher in preparing the children for the activity.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate? (clear, simple)

The test booklets contain no more than two questions per page, each is distinct from the brown background of the page.

Are the printing and/or illustrations of the appropriate style, size, and clarity? (clear, simple)

Children of several racial backgrounds are portrayed. Several of the animals are difficult to identify. (The "tiger" on page 21 of CIRCUS 13 for example.)

Are the length of the test and the time pressures appropriate?

Most tests are untimed. The teacher is encouraged to use her own judgement.

How are answers recorded?

For most instruments, the child marks through the picture corresponding to the correct answer. The teacher records some of the answers.

What equipment is needed?

Pencil.

Is it readily available?

Yes.

How much maintenance does it require?

None. A cassette recorder is needed with one test.

RELIABILITY, REPLICABILITY, REFINEMENT

The alpha reliability, an estimate of the internal consistency of a test, is reported for all measures for which it can be computed.

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

Not given.

What is the correlation between two comparable or alternate forms of a test taken by examinees? (equivalence reliability should be at least .70)

Not given.

Are procedures for replicating the statistical procedures clearly described?

The Technical Manual must be ordered in addition to the specimen set to obtain this information.

How are converted scores reported?

None are reported.

TEACHING FEEDBACK

What is the norm range? (more than 3 years in age, 3 years of education, and a variety of ability levels)

The children were 4 and 5 and attended a variety of pre-primary education centers.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?

The norms are current. They over-represent children in the Northeast and children in cities of less than 50,000 and under-represent children in the Southeast, children in cities over 50,000, and black children.

Are procedures for obtaining a new score and converting the score (where applicable) clearly explained and easy to execute?

Yes, the scores can be located in a table. Each range of scores corresponds to a sentence report which describes both the general level of competence and areas which need improvements.

Is special training required to interpret the test scores?

No.

Can this test be directly, clearly, immediately interpreted?

Yes.

Does the test provide guidelines for making decisions from a score or score interpretation?

The sentence reports contains suggestions for areas which need improvement.

USEABILITY

Is this test administered to individuals or groups?

Usually groups of up to ten.

How much time does administration take?

Untimed.

Does the manual contain a discussion of the purpose, use, and limitation of the test?

Extensive information is provided about the purpose and use of the tests.

Does the manual contain clear administration and scoring directions?

Many excellent suggestions are made facilitating test administration; scoring is straightforward.

Does the manual provide a description of test development, analysis procedures, and normative groups?

A comprehensive discussion of the test development and normative groups are provided. More description of the analysis procedures would be helpful.

Can the test be administered by the teaching staff without extensive practice?
Yes.

How much time is necessary to prepare for the test administration?
The teacher should be familiar with the test items.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive statistics, and reliabilities?
Not applicable.

SUMMARY

Educational Testing Service designed CIRCUS as a comprehensive array of assessment devices. A large, though slightly unrepresentative sample provides a significant base on which interpretation of scores can be made. The "sentence report" format for reporting scores provides information about the child's strengths and areas that need some additional work. The major focus of most of the instruments are activities generally classified as "readiness." If a program were looking for an instrument to assess "readiness," CIRCUS should be included in the list of potential instruments.

Title: Comprehensive Development Scale (Project MEMPHIS)

Authors: Alton Quick, Thomas Little, and Ann Campbell

Publisher: Lear Siegler/Fearon Publishers
6 Davis Drive
Belmont, CA. 94002

Description:

The MEMPHIS Comprehensive Developmental Scale is a part of Project MEMPHIS' three step system for developmental-educational evaluation and educational programming for the preschool handicapped child. The instrument may be used separately from the other parts of the system. The Comprehensive Developmental Scale is described as a part of a longer publication describing the entire project. No information is presented which describes how the instrument was developed. No data is presented converning the reliability or validity of the instruments. While such information may be listed in another publication, no reference is made to the availability of the information. The Scale is divided into five developmental areas: (1) personal-social, (2) gross motor, (3) fine motor, (4) language, and (5) perceptive-cognitive. The authors emphasize that the scores from this scale should be thought of as rough estimates of the child's developmental levels for purposes of gaining information about deficiencies and needed educational treatment emphasis. (The Scale should not be used for educational of developmental diagnoses.) A set of three forms and one guide costs $1.50. (1974)

VALIDITY

What does this test measure?

This scale assesses developmental status in these areas: fine motor skills, gross motor skills, personal-social skills, language skills and perceptive-cognitive skills.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

At least two items are included for each three month period--the items seem to encompass major developmental milestones.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

No sources of the items or procedures for assigning the items to particular age categories are indicated.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

Most items can be observed in usual classroom activities.

Are instructions to the child about what he is to do and why clear and suitable in tone?

None are indicated.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)
Yes.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?
Yes.

Are the length of the test and the time pressures appropriate?
The Comprehensive Developmental Scale is untimed.

How are answered recorded?
The examiner records the answers on a scoring sheet.

What equipment is needed?
None is specified but some items require commonly available materials.

Is it readily available?
Yes.

How much maintenance does it require?
Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)
No indication is given of any reliability or validity studies.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)
No alternate forms are available.

Are procedures for replicating statistical procedures clearly described?
No statistical analyses are indicated.

How are converted scores reported?
A developmental profile can be constructed.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)
The population served by Project MEMPHIS is handicapped preschoolers. No indication is given as to how the age levels on the Scale were determined.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?
No norms are available.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?

Yes.

Is special training required to interpret the test scores?

Yes. The scores are considered to be rough estimates of the child's developmental level. The interpretation of the score in terms of educational programming requires special training.

Can this test be directly, clearly, immediately interpreted?

No. It provides a starting point for educational planning.

Does the test provide guidelines for making decisions from a score or score interpretation?

Project MEMPHIS, a book which describes not only the Comprehensive Developmental Scale but also the other components of the Project, contains descriptions of the utilization of the Scale in the context of the other components. In relation to the other components of the system, guidelines for making decisions are indicated. Sample lesson plans are also included.

USEABILITY

Is this test administered to individuals or groups?

Individuals.

How much time does administration take?

Untimed, no indication is given of the expected length.

Does the manual contain a discussion of the purpose, use and limitation of the test?

Yes, general information is provided.

Does the manual contain clear administration and scoring directions?

The scoring directions are clear but only meager information is provided about administration.

Does the manual provide a description of test development, analysis procedures, and normative groups?

No indication is given of how the test was developed, how the items were selected or ordered, or the groups which might have been used to provide normative information.

Can the test be administered by the teaching staff without extensive practice?

Once administration procedures were established, the Scale could be administered without extensive practice.

How much time is necessary to prepare for the test administration?

The amount of time is difficult to determine in the absence of a description of procedures for administration.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

If the Comprehensive Developmental Scale is used as the authors suggest to obtain rough estimates of the child's developmental levels for the purpose of gaining information for educational planning based on the child's strengths and weaknesses, then the Scale may provide some useful information if used by a trained examiner. Since evidence of reliability and validity are lacking, diagnoses made on the basis of this Scale are unwarrented.

Title: Comprehensive Identification Process

Author: R. Reid Zehrbach

Publisher: Scholastic Testing Service
480 Meyer Road
Bensenville, Illinois 60106

Description:

The Comprehensive Identification Process was designed to help identify preschool children who may be eligible for special services prior to school entry. CIP identifies for review, referral, or evaluation children who may have problems in these areas: fine motor, gross motor, cognitive-verbal, speech and expressive language, hearing, vision, socio-affective behavior, and medical history. Children attend three stations for various sections of the CIP. At the same time, parents provide additional information about the child on an interview form. Paraprofessionals can administer CIP with 2 to 6 hours training. The age range is from 2½ to 5½ years. The kit with all manuals and materials for 35 screenings is $59.95. (1975)

VALIDITY

What does this test measure?

CIP screens the speech and language, gross and fine motor, socio-affective, and cognitive behavior of children as well as their vision.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of time)

As a screening instrument, CIP seems to have an appropriate number and complexity of items. The author indicates that items were selected through a study of the research in the area including the abstracted descriptions of the test items contained in the Developmental Guidelines by Karnes, et al. The criterion for item selection are listed. The claim of validity would be strengthened if more references were cited.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journal)

The author's presentation of results of the initial use of CIP with 762 children notes that only 7 children (.9 percent) were passed at the final level of the screen. Specific information is presented about the number of referrals and passes at each level of the process. The same specificity of data is not presented with the three replication studies.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

Yes, the children will probably enjoy the activities.

Are instructions to the child about what he is to do and why clear and suitable in tone?

Yes.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)

Some of the stylized pictures such as the purse and the ring are difficult to interpret.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

The item packaging is particularly effective. All materials are stored in a plastic "briefcase." Some of the colors in the pictures are very strong and might be distracting.

Are the length of the test and the time pressures appropriate?

Yes. The CIP takes 30-40 minutes.

How are answers recorded?

Each screener records a "pass," "fail," or "refused" on the Record Form.

What equipment is needed?

Yes, most materials are supplied however.

Is it readily available?

Yes.

How much maintenance does it require?

Minimial. The CIP seems very durable.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

A "presentation brief" from the publisher indicates that the inter-rater reliability are within the .90-.95 level by item. This information is not reported in the manual. Of the 58 children in the initial study who were rescreened, 26 passed. The author notes the selection criteria will effect the number of passes.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)

None are indicated.

Are procedures for replicating statistical procedures clearly described?

Yes.

How are converted scores reported?

Three scores are generated: "P" for pass and normal for age, "R" for refer for specific work, review the questionable results, or rescreen if the results are in doubt and "E" for evaluate.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)

The age range is 3 to 5 years. No indication of the identification of ability level is given.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?

The manual contains no breakdown by demographic information. The final form of the CIP was administered to children who are essentially all Caucasian and from small mid-western communities and rural areas. A "presentation brief" from the publisher notes that the original sample was cross-cultural and included caucasian and black children. This information is not included in the manual.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?

Criterion levels for each item, with the exception of speech and language, are specified.

Is special training required to interpret the test scores?

Yes. A professional would be required to review the child's folder. Many cases are immediately obvious.

Can this test be directly, clearly, immediately interpreted?

Yes.

Does the test provide guidelines for making decisions from a score or score interpretation?

Yes, specific levels are set for pass, rescreen, and evaluate. Professional judgement is required in some cases, however.

USEABILITY

Is this test administered to individuals or groups?

Individuals. Parents also provide information on a Parent Interview Form.

How much time does administration take?

30-40 minutes.

Does the manual contain a discussion of the purpose, use and limitation of the test?

The manual contains extensive discussion of the screening procedure including directions for organizing the entire process. The author notes that professional interpretation may be necessary in many cases.

Does the manual contain clear administration and scoring directions?

Yes, however, including a copy of the records form in the manual would be helpful.

Does the manual provide a description of test development, analysis procedures, and normative groups?

Details of the test development are included. Unfortunately only limited information is presented about the normative group.

Can the test be administered by the teaching staff without extensive practice?

The author suggests that a 2 to 6 hour workshop be held in order to train the staff. The amount of time required will depend on the background of the screeners.

How much time is necessary to prepare for the test administration?
Approximately one hour should be allowed to set up the screening site.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

CIP is designed to identify children who are within the lowest 10-15% of the general population within their age group. The screening procedure seems to yield accurate results. In the initial study, only a part of a percent passed at the final level. One interpretation of this finding is that CIP is probably locating most of the children who need services. An administrator would be more confident in using the CIP is more adequate information about the norming group were supplied, particularly in light of the publisher's statement that the second sample was cross cultural and included caucasian, black, and oriental children.

Title: Cooperative Preschool Inventory, Revised Edition

Author: Bettye Caldwell

Publisher: Cooperative Tests and Services
Educational Testing Service
Rosendale Road
Princeton, New Jersey 08540

Description:

This Inventory was designed as a screening procedure for areas usually associated with school radiness. The emphasis is on verbal items of the "information" type with several items requiring ability to label quantities, identify serial position, perceive shapes, and perform visual-motor drawing tasks. The age range is 3 years to 6 years. The cost is $2.50 for 20 tests, $1.00 for a technical report and $1.00 for a manual. (1970)

VALIDITY

What does this test measure?
The major focus is comprehension and information.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)
The heavy emphasis on verbal "information" items provides little information about other important language areas.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)
The Technical Report lists the percent of children at each age who pass a particular item. At least a quarter of the items are only minimially discriminative. No indication is given of the correlation of this instrument with other people.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?
Yes.

Are instructions to the child about what he is to do and why clear and suitable in tone?
Yes.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)
Yes.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?
Appropriate but not very stimulating to the child.

Are the length of the test and the time pressures appropriate?
The test is untimed but should take approximately 15 minutes.

How are answers recorded?
The examiner marks a scoring sheet. Space is also provided for the child's drawings.

What equipment is needed?
Pencil, manual, recording forms and several materials which must be obtained locally. Directions are included for making boxes used with some tasks.

Is it readily available?
Yes.

How much maintenance does it require?
Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)
No indication of test-retest reliability is given. The internal consistency of the test was measured using split-half correlations. Many items measuring similar concepts contribute to the high correlations which average .92.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)
No.

Are procedures for replicating statistical procedures clearly described?
The Technical Report contains a discussion of many procedures used in the instrument. Percentile conversion charts are reported for middle and lower class groups. No specific criteria for placing children in the norm groups is indicated.

How are converted scores reported?
Converted scores are not reported.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)
The age range includes children 3 years to 6 years 6 months.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?
The sample included 1531 Head Start children in 150 Head Start classes across the United States in the fall of 1969.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?
Not applicable.

Is special training required to interpret the test scores?
No indication is given of the interpretation of the results.

Can this test be directly, clearly, immediately interpreted?
No indication is given.

Does the test provide guidelines for making decisions from a score or score interpretation?
None are indicated.

USEABILITY

Is this test administered to individuals or groups?
Individuals.

How much time does administration take?
Approximately 15 minutes.

Does the manual contain a discussion of the purpose, use and limitation of the test?
The Inventory was originally developed to measure changes in children as a result of Head Start intervention.

Does the manual contain clear administration and scoring directions?
Directions are clear and complete.

Does the manual provide a description of test development, analysis procedures, and normative groups?
Extensive description of the test development is included. Analysis procedures is also indicated.

Can the test be administered by the teaching staff without extensive practice?
Yes.

How much time is necessary to prepare for the test administration?
Minimal.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

The Cooperative Preschool Inventory lacks sufficient information on validity and reliability. It measures only a limited aspect of school readiness and contains no information which would help in the interpretation of the information from the Inventory.

Title: Denver Developmental Screening Test

Authors: W. Frankenburg and J. B. Dodds

Publisher: Ladoca Project and Publishing Foundation, Inc.
East 51st Avenue and Lincoln
Denver, Colorado 80216

Description:

The Denver Developmental Screening Test is designed to assist in the early detection of children with serious developmental delays. The authors note that the DDST is not a diagnostic tool. Rather, it alerts the examiner to the presence of a developmental problem in the following areas which need further investigation: gross motor, fine motor-adaptive (use of hands, ability to solve non-verbal problems), language (ability to hear and talk), and personal-social (self-care tasks, ability to relate to others). Usually only about 20 of the 105 tasks are administered to a child. The test is individually administered. Task norms indicate the age at which 25%, 50%, 75% and 90% of the total sample passed the items. The age range includes 2 weeks to 6 years. (1969)

VALIDITY

What does this test measure?
This test assesses motor, language, and social development.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)
The items were selected from other scales and then field tested to determine the final pool of items. An appropriate number and complexity of items seems to be included.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)
Buros' 7th lists 3 studies in which the DDST is used. The authors note that the DDST reveals only minor differences by sex, but children whose fathers are in white collar occupations tend to acquire language skills earlier than children whose fathers are in blue color occupations.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?
Yes.

Are instructions to the child about what he is to do and why clear and suitable in tone?
There is no indication of instructions to be given to the child.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)
The figures which the child must copy or identify from the back of the test are small and appear in the "Directions" rather than separately.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

The scoring sheet seems complicated but the directions for scoring are clear and complete.

Are the length of the test and the time pressures appropriate?

The DDST is untimed; the child is given three trials to perform each item.

How are answers recorded?

The examiner marks the answer sheet.

What equipment is needed?

Yes.

Is it really available?

Yes, a small zipper bag of the following standardized materials is provided: red wool, raisins, rattle with narrow handle, eight one-inch cubes, bottle, bell, ball, and pencil.

How much maintenance does it require?

Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

A sample of twenty children ranging in age from 2 months to 5 years were tested twice by the same examiner at an interval of one week. The reliability was .95. However, more than one examiner's test-retest scores will be needed to adequately establish this reliability. Reliability among testers was reported as an average percent of agreement on items passed or failed.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)

Not applicable.

Are procedures for replicating statistical procedures clearly described?

General procedures for application of statistical procedures are reported.

How are converted scores reported?

Scores are reported as "pass," "fail," "refused," or "no opportunity." If the child fails an item below his age line, the child is considered to have a delay in the area.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)

The norm range included children from 1 to 72 months of all SES categories. No indication is given that other measures of ability were administered to the norming population. The authors mention a "preliminary study" of validity using 18 children from 4 to 68 months; these children are not indicated as a part of the norming group.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?

No date is given for standardization but it is assumed to be before the publication of an article on the DDST in 1967. A serious defect of the DDST is the lack of representativeness of the norming population. The geographic area is restricted to Denver. Higher SES groups are over-represented. Mothers consented to having their children tested. As the authors point out, all these factors contribute to the inappropriateness of the Denver for all children.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?

Yes, pictures and explicit directions are included.

Is special training required to interpret the test scores?

Specific, brief procedures are provided for interpreting the test as "abnormal," "questionable," and "normal."

Can this test be directly, clearly, immediately interpreted?

The scores can be categorized as described above. The child with an "abnormal" or "questionable" should be rechecked in several weeks and referred to a doctor if the score is the same.

Does the test provide guidelines for making decisions from a score or score interpretation?

Yes, on a very limited basis. Case studies are also provided.

USEABILITY

Is this test administered to individuals or groups?

Individuals.

How much time does administration take?

Untimed.

Does the manual contain a discussion of the purpose, use and limitation of the test?

Yes, the authors are to be commended for their emphasis on the DDST as a screening instrument.

Does the manual contain clear administration and scoring directions?

Yes, examples are also provided to aid in interpretation.

Does the manual provide a description of test development, analysis procedure, and normative groups?

Yes, adequate information is provided.

Can the test be administered by the teaching staff without extensive practice?

Training would seem to be required for the examiner to accurately assess the behaviors. Specific instructions are included for the administration and scoring of each item.

How much time is necessary to prepare for the test administration?

The examiner must be thoroughly familiar with the items.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

If used as a screening instrument as the authors suggest, then the DDST can provide the trained examiner with some useful insight on the development of a particular child. The most serious problem with the DDST is the lack of representativeness in the norming population of geographic, SES, and racial differences. This neglect is serious if the child being screened is from a population different from the norming population.

Title: Developmental Progress Scale

Author: Robert D. Boyd

Publisher: None is indicated. (Review copy was obtained from Becky Cook.)

Description:

The Developmental Progress Scale is designed to portray visually the child's level of behavior in Motor Skills, Communication Skills, Self-Sufficiency Skills and to identify those skills which are below the child's general level of functioning. Part of the information about the child is obtained during an interview with the parent. Additional information is obtained during the individual administration of the Scale. The Developmental Progress Scale is very similar to the Denver Developmental Screening test with a larger number of items included in the age levels from 5 to 8 than is included in the DDST. The age range is birth to 8 years. (1973)

VALIDITY

What does this test measure?

This test measures motor skills, language skills, self-help skills. The items were selected on the basis of research evidence that indicated the particular behavior or skill under consideration typically emerged in 60% to 70% of the children at that age level. Since no references are listed, it is difficult to validate this statement.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

The Scale has an insufficient number of items in the lower age levels. Some of the items such as "opens closed doors" (12-18 months) and "rides tricycle" (2-3 years) seem to be on the low side.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

None is indicated.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

For the most part. However, the author suggests that a reading test should be administered to the child. The child is expected to score at the .5 level between the 5th and 6th year.

Are the instructions to the child about what he is to do and why clear and suitable in tone?

Yes.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)

The Scale is printed on one page to facilitate scoring. It is easy to read.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?
The copy seen by this reviewer was xeroxed; insufficient information is available to answer this question.

Are the length of the test and the time pressures appropriate?
The Scale is untimed.

How are answered recorded?
The examiner marks the score sheet.

What equipment is needed?
A few objects such as a bottle and ball will need to be assembled locally.

Is it readily available?
Yes.

How much maintenance does it require?
Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)
No information, other than a general statement that items were drawn from the literature, is presented about the reliability of this Scale.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)
Alternate forms have not been developed.

Are procedures for replicating statistical procedures clearly described?
The procedures are not indicated.

How are converted scores reported?
All items assessed for a particular child are marked on the Scale. At the end of the testing, a profile of the child's strengths and weaknesses can be determined by scanning the items passed on the Scale.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)
No norms are indicated.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?
No norms are indicated.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?
The Profile is constructed as the test is scored.

Is special training required to interpret the test scores?
A teacher with skills in task analysis could interpret the Scale.

Can this test be directly, clearly, immediately interpreted?
No, many judgements are required about the meaning of particular scores.

Does the test provide guidelines for making decisions from a score or score interpretation?
Some general suggestions are provided.

USEABILITY

Is this test administered to individuals or groups?
Individuals.

How much time does administration take?
No indication is given.

Does the manual contain a discussion of the purpose, use and limitation of the test?
The author discusses developmental scales in general and makes some suggestions regarding the use of the test. Since the Developmental Progress Scale is very similar to the Denver, the author should provide a more complete discussion of the purpose.

Does the manual contain clear administration and scoring directions?
Yes. The administration and scoring directions are presented together to facilitate scoring. In addition simplified administration procedures which group the items are included.

Does the manual provide a description of test development, analysis procedures, and normative groups?
None are indicated.

Can the test be administered by the teaching staff without extensive practice?
Possibly. The author states that the items do not need to be administered in a specified order. A skilled examiner seems to be required.

How much time is necessary to prepare for the test administration?
Minimal.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, or reliabilities?
Not applicable.

SUMMARY

The Developmental Progress Scale is very similar in terms of format and items to the Denver Developmental Screening Test. It includes 45 more items in the upper range than does the DDST. However, no specific information about the validity of the Scale is offered. Without specific references to sources from which the items were drawn, claims of validity are unwarranted. No evidence of reliability is presented.

Title: Developmental Screening Questionnaire

Author: Elizabeth Y. Sharp

Publisher: Model Preschool for Handicapped Indian Children
Department of Special Education
University of Arizona
Tuscon, Arizona 85721

Description:

The Developmental Screening Questionnaire was designed for use by Head Start programs on Indian Reservations in Arizona. It is used to identity children who may have problems in physical development, self-care, language development, and social and emotional development. Identified children should be referred for evaluation. The age range is 3 years to 5 years. ·(1974)

VALIDITY

What does this test measure?
This instrument measures physical development, language development, social and emotional development.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)
This instrument provides only a minimum sampling of behavior. It is intended to provide only a general screening of children.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)
None is presented.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?
Most items can be observed in the context of usual classroom activities. The tasks do not specify criterion levels so variability among examiners is to be expected.

Are instructions to the child about what he is to do and why clear and suitable in tone?
None are indicated.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)
Yes.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?
Yes.

Are the length of the test and the time pressures appropriate?
The Questionnaire is untimed.

How are answers recorded?
The answers are recorded on a scoring sheet.

What equipment is needed?
Materials commonly found in the classroom are used for some items.

Is it readily available?
Yes.

How much maintenance does it require?
Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)
None are indicated.

What is the correlation between two comparable or alternate forms or the test? (equivalence reliability should be at least .70)
None are indicated.

Are procedures for replicating statistical procedures clearly described?
None are indicated.

How are converted scores reported?
Raw scores are used.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)
None are indicated.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?
None are indicated.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?
Raw scores are used.

Is special training required to interpret the test scores?
No. If the child does not pass a specified number of items, he should be referred for evaluation. Training in observation is required to obtain accurate scores.

Can this test be directly, clearly, immediately interpreted?
Yes.

Does the test provide guidelines for making decisions from a score or score interpretation?
The child should be referred for evaluation if a certain number of items are missed.

USEABILITY

Is this test administered to individuals or groups?
Individuals, though specific children could be observed as they participate in group activities.

How much time does administration take?
Unspecified.

Does the manual contain a discussion of the purpose, use and limitation of the test?
None are indicated.

Does the manual contain clear administration and scoring directions?
The scoring directions are clear but the administration directions are unspecified.

Does the manual provide a description of test development, analysis procedures, and normative groups?
No descriptions are provided.

Can the test be administered by the teaching staff without extensive practice?
The staff should have observation skills.

How much time is necessary to prepare for the test administration?
Unspecified.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

The lack of any information on the validity and reliability of this Questionnaire, along with the lack of specification of the tasks, are detriments to its apparent simplicity. This Questionnaire should be used with caution.

Title: DIAL (Developmental Indicators for the Assessment of Learning)

Authors: Carol D. Mardell and Dorthea S. Galdenburg

Publisher: DIAL, Inc.
Box 911
Highland Park, Illinois 60035

Description:

The DIAL is an individually administered screening instrument for gross motor, fine motor, concepts, and communication skills. Four stations are set up; the child moves from station to station after having an opportunity to play with clay. Paraprofessionals may be trained to administer the DIAL under the supervision of a coordinator. The age range is from 2½ to 5½ years. (1972)

VALIDITY

What does this test measure?

This instrument provides information about a child's functioning in motor skills, language, and cognitive skills.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

The tasks were selected to reflect a conceptual model to include the six major areas of a child's functioning. Both a task analysis and a description graphic representation are included.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journal)

The authors list the norm dates and references for each item. The authors also listed the panel of consultants which reviewed the DIAL. In order to assess concurrent validity, a diagnostic team administered a battery of standaridzed tests to children that had been screened with the DIAL. There was 85.3% agreement between the two teams. Predictive validity with the Iowa Test of Basic Skills, the Metropolitan Achievement Test, Stanford Achievement Test is reported but the results of the calculation of the Multiple R are difficult to interpret. Item analyses are also included for each item.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

Yes. The children probably particularly enjoy talking about their photographs.

Are instructions to the child about what he is to do and why clear and suitable in tone?

Yes.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)

The illustrations are clear and simple.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

Several subtests utilize two disks for presentation of the information to the child. Pictures or shapes are on the bottom disk. The top disk has a circular cut out the size of one picture. The child is not distracted by the other pictures. All the DIAL materials are stored in a thick cardboard barrel approximately 2 feet high. The addition of handles to the barrel would make it easier to transport.

Are the length of the test and the time pressures appropriate?

The DIAL takes 25 to 30 minutes per child. Two of the subtests are timed but the limits seem reasonable.

How are answered recorded?

The screener marks a score sheet. Cues are provided on the score sheet to aid in scoring.

What equipment is needed?

Most equipment is provided. However, a watch with a second hand and a Polaroid camera are also needed.

Is it readily available?

No.

How much maintenance does it require?

Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

A group of children who had been tested with the DIAL in the spring of 1972 were retested a year later. Half of the children had been in the upper 90% and half had been in the lower 10%. The two groups remained very distinct.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)

None are indicated.

Are procedures for replicating statistical procedures clearly described?

No, some of the procedures are difficult to follow which makes interpretation difficult.

How are converted scores reported?

Raw scores are converted to scaled scores by simple multiplication.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)

On a population of 3100 children, one hundred children were randomly selected for each month of age from 36 to 66 months and controlled for sex to establish separate norms for males and females.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?

Though extensive description is offered of the standardization sample, however, no information about the norming sample is provided.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?

Yes.

Is special training required to interpret the test scores?

Yes. Professional judgement is required for some scores.

Can this test be directly, clearly, immediately interpreted?

Yes.

Does the test provide guidelines for making decisions from a score or score interpretation?

Specific scores are set for "O.K.," "REDIAL," and "FOLLOW-UP." However, many factors may influence a child's score.

USEABILITY

Is this test administered to individuals or groups.

The DIAL is a team administered individual screening instrument.

How much time does administration take?

25-30 minutes.

Does the manual contain a discussion of the purpose, use and limitation of the test?

Yes, the authors emphasize that the DIAL is to be used for screening only. A helpful feature of the manual is a list of additional tests which could be used for further evaluation.

Does the manual contain clear administration and scoring directions?

Yes, and criteria are specified for the items.

Does the manual provide a description of test development, analysis procedures, and normative groups?

Extensive data is provided of the test development. The inclusion of references for specific items is helpful. The analysis procedures should be studied carefully; much information seems to be presented. Normative groups need to be defined.

Can the test be administered by the teaching staff without extensive practice?
Some training is required before the DIAL can be administered.

How much time is necessary to prepare for the test administration?
About an hour should be allowed for setting up the screening site.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

The DIAL provides a procedure for screening gross motor, fine motor, communication, and concepts (cognitive) skills. Efforts have been made by the author to establish the validity and reliability of the items. If the norm group, however, is similar to the standardization sample, it may not be representative of children outside of Illinois. The DIAL does not screen for social or self-help skills.

Title: Koontz Child Developmental Program: Training Activities for the First 48 Months

Author: Charles W. Koontz

Publisher: Western Psychological Services
12031 Wilshire Boulevard
Los Angeles, California 90025

Description:

The Koontz Child Development Program was designed to provide a tool for developmental assessment which incorporates a method of plotting progress and a method of planning suitable activities. The child's behavior is observed in a normal setting and then is matched to a list of graded performance items in four areas: gross motor, social, language, and fine motor. The Koontz Program is intended to be used by teachers to observe a child's behavior and then plan appropriate activities. It is intended for children from birth to 48 months. (1974)

VALIDITY

What does this test measure?
This test measures the child's motor, language, and social behavior.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

A total of 550 performance items can be measured, 25 performance items for each of the 22 age level. Indication is given of items that should be modified for a hearing or visually impaired child. Criterion for evaluation are also included.

What empirical evidence exists that would give an administrator confidence in its usage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

The author notes that items were selected from existing scales such as the Bayley or from experts in the field. However, only 4 references are listed. The author did specify the criteria on which the items were selected. The author presents validity information on thirty-four severely retarded and moderately retarded children in three different settings who participated in the program. Pre-post scores are reported in addition to difference scores. No control groups are reported nor is control for variables such as specific kind of handicapping condition or way of determining IQ indicated.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

The evaluation items are generally appropriate. However, some of the Training Activities, such as "continue manipulation of the baby constantly" (2 months) would seem to encourage the child to not respond due to fatigue.

Are instructions to the child about what he is to do and why clear and suitable in tone?

None are indicated.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)

Yes.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

The Program is printed on heavy paper and packaged in a spiral notebook. The evaluation and training activities for each level are printed on separate strips. The teacher observes the child and locates the level at which the child is performing. The division of the page into separate strips facilitates observing the child whose behavior in different areas is at different levels.

Are the length of the test and the time pressures appropriate?

It is suggested that the child be observed for at least one week before the evaluation begins. No indication as to how much observation time during the week would be required.

How are the answers recorded?

The evaluator marks a record card.

What equipment is needed?

Yes. The child is to be observed in his regular environment.

Is it readily available?

Yes.

How much maintenance does it require?

Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

Twenty-three children were evaluated by their teachers and then evaluated after 2 to 6 weeks by another teacher. Out of 92 possible evaluations, 72 were in complete agreement, a reliability coefficient of .78.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)

No alternate forms are indicated.

Are procedures for replicating statistical procedures clearly described?

Yes, the procedures are described.

How are converted scores reported?

No converted scores are reported.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)

The author does not specify how the age levels of performance items were determined. No indication is given of norm ranges.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?

No indication of any norming groups are given.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?

No converted scores are reported.

Is special training required to interpret the test scores?

No. The teacher can plan activities on the basis of the scores. Some activities are suggested for each item. The author cautions that though the items are in developmental sequence, they may be at age levels different from other sources.

Can this test be directly, clearly, immediately interpreted?

Yes.

Does the test provide guidelines for making decisions from a score or score interpretation?

Yes, specific training activities are included. The activities should be reviewed carefully to determine the appropriateness of the activities.

USEABILITY

Is this test administered to individuals or groups?

Individuals are observed.

How much time does administration take?

The child should be observed for at least a week.

Does the manual contain a discussion of the purpose, use and limitation of the test?

Yes, the brief description seems adequate.

Does the manual contain clear administration and scoring directions?

Yes. The criteria for each item is also specified.

Does the manual provide a description of test development, analysis procedures, and normative groups?

Criteria for the selection of items are specified. Unfortunately, the sources of the items are not specified.

Can the test be administered by the teaching staff without extensive practice?

Yes. The staff must be able to observe accurately, however.

How much time is necessary to prepare for the test administration?

Minimal.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

The Koontz Program is organized in a format that facilitates observation of the child. The training activities indicated for each task should be evaluated carefully by potential users because the child may be unduly fatigued by excessive stimulation. The reliability and validity information are reported only for a small number of retarded children. Generalizations to normal children would be difficult to make. This Program is intended for children up to 48 months. It would not be appropriate for assessment of normal four year olds due to the low ceiling of the Program.

Title: Learning Accomplishment Profile

Author: Anne R. Sanford

Publisher: Kaplan School Supply Corp.
600 Jonestown Road
Winston-Salem, N. C. 21703

Description:

The Learning Accomplishment Profile was designed to provide the teacher of young handicapped children with a criterion-referenced assessment of the child's gross motor, fine motor, social, self-help, cognitive and language skills. The first section of the LAP contains a hierarchy of developmental data for assessing the child's existing skills; the second section provides guidance in sequencing skill development, and the third section provides 44 weeks of curriculum units. The LAP is individually administered by a teacher although many items can be observed in regular classroom activities. The age range is 1 month to 72 months. The cost is $1.00 per manual and $1.00 per LAP. (Cost figures may be slightly dated.) (1974)

VALIDITY

What does this test measure?

The LAP provides information about the child's motor, language, cognitive, social, and self-help skills.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

The number of items provided at each level varies so that the level of a child's performance is sometimes difficult to compute. In some cases, a behavior, such as "gives use of object" is assessed in two different categories. Language Development and Cognitive has several overlapping items.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

The author cites the source of each task and, for many tasks, includes comments from the original sources. No validity data is reported at this time. A Field Test Edition of a modification of the LAP is currently being tested.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

Yes.

Are instructions to the child about what he is to do and why clear and suitable in tone?

Instructions to the child are not indicated.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)

Yes.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

The separation of the Developmental Scale from the Task Hierarchy can be confusing. Once the child has missed four of five items in an area, those items indicate potential objectives for the child. The teacher can then refer to the task analysis section for help in determining the specific level of the task the child has not mastered. The inexperienced teacher might attempt to teach to the task rather than determine which components of the task the child has mastered.

Are the length of the test and the time pressures appropriate?

The LAP would probably be used throughout the year.

How are answers recorded?

The teacher marks the appropriate box on one of the 42 pages of items.

What equipment is needed?

Yes. Equipment usually found in the classroom is needed.

Is it readily available?

Usually. In some centers, the collection of pictures to assess some of the language and cognitive items is a bit difficult due to lack of resources.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

No correlations are reported.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)

No correlations are reported.

Are procedures for replicating statistical procedures clearly described?

No procedures are reported.

How are converted scores reported?

Scores can be displayed on a Profile. No indication is given that the cut-off level of errors on 4 out of 5 items actually corresponds to a particular developmental level. A child's rate of development can be computed. The change in the rate of pupil development after intervention can be measured. No attempt is made to explain the cause of increase.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)

No norm range is indicated.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?
No norms are indicated.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?
Insufficient information is provided for constructing the Profile, particularly for a child that has not completed all the tasks at a particular level.

Is special training required to interpret the test scores?
No. Skills will be required to implement a program based on the LAP, particularly if the LAP is conducted with several children.

Can this test be directly, clearly, immediately interpreted?
Yes.

Does the test provide guidelines for making decisions from a score or score interpretation?
Yes. The task hierarchy and curriculum sections contain information that would be useful to the teacher. A separate curriculum is also available.

USEABILITY

Is this test administered to individuals or groups?
It can be administered to individuals or groups.

How much time does administration take?
No amount of time is specified. The LAP can be used as an ongoing assessment procedure.

Does the manual contain a discussion of the purpose, use and limitation of the test?
The manual describes criterion-referenced assessment, and a variety of uses of the LAP. Limitations are briefly discussed.

Does the manual contain clear administration and scoring directions?
The administration descriptions are general. The marking procedures are described and criterion levels are provided for many, but not all, of the items.

Does the manual provide a description of test development, analysis procedures, and normative groups?
None are indicated.

Can the test be administered by the teaching staff without extensive practice?
The teachers need to be skilled in observation, particularly if more than one child is to be observed at the same time.

How much time is necessary to prepare for the test administration?
The LAP is often used as an ongoing assessment procedure. The preparation time will vary.

RETEST POTENTIAL

Are alternate forms of the test available.
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

The Learning Accomplishment Profile provides information to aid the child in refining motor, language, cognitive, social, and self-help skills. Generally, the items provide the teacher with abundant opportunities to observe a child's functioning. It must be noted that some items are redundant and some lack specification of criteria. Though information is provided on task hierarchies and some curriculum units are suggested, training would be needed to transform the information from the LAP into a program for the child. It is hoped that the field test being constructed on the LAP will yield information on the validity and reliability of the LAP. A review of the literature regarding the computation of the rate of development is in order. At this time, since the intervals between items are not uniform, the method of calculating the rate of development seems unwarranted.

Title: Peabody Picture Vocabulary Test

Author: Lloyd M. Dunn

Publisher: American Guidance Service, Inc.
Publishers Building
Circle Pines, Minnesota 55014

Description:

The PPVT is designed to measure receptive vocabulary through measuring the child's hearing vocabulary as the child points to the picture that corresponds to the correct answer. It attempts to predict school success, particularly in areas which call for verbal intelligence. The pictures were selected from an initial pool of 3,885 words from Webster's New Collegiate Dictionary. The 154 page picture booklet is used with both forms. A manual and individual test record form are also included. The PPVT is individually administered to children between the ages of 2 years 6 months and 18 years. The test materials may be purchased for $10.00 which include 25 record forms. (1959-1970)

VALIDITY

What does this test measure?

The PPVT measures hearing vocabulary with the inference often made that verbal intelligence is also being measured. In the validity studies reported in the manual which correlate PPVT with the Stanford-Binet, the WISC, SCAT, STEP and other standardized measures, the range of correlation coefficients is substantial. It must be noted that several of the studies included individuals from restricted environments. Restricted IQ ranges tend to lower the correlation between the PPVT and measures of general intelligence which tap a wider range of behaviors.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

The PPVT contains a large pool of items which were chosen according to specified criteria. It provides a reasonable assessment of receptive vocabulary but does not measure all aspects of verbal intelligence.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

The PPVT is a widely used clinical tool. The manual lists 44 studies which used the instrument. The coefficients of the several types of reliability range from .54 to .97. Buros 6th and 7th also contains an extensive reference listing.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

Yes.

Are instructions to the child about what he is to do and why clear and suitable in tone?

Yes. Separate instructions are included for those under eight and those over eight.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)

The pictures are clear, simple line drawings. Minority groups, particularly blacks, are significantly under-represented.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

Yes.

Are the length of the test and the time pressures appropriate?

Yes.

How are answers recorded?

The examiner records the number on the scoring sheet. Scoring is both fast and reliable.

What equipment is needed?

Pencil, picture booklets, record form, manual.

Is it readily available?

Yes.

How much maintenance does it require?

Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

For the studies reported in the manual, the mean correlation is .74 with a range of .88 to .54.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)

For the studies reported in the manual, the mean correlation is .87 with a range of .75 to .97. These correlations are comparable to those found in the standardization population.

Are procedures for replicating statistical procedures clearly described?

The discussion of the test standardization is comprehensive as is the discussion of the procedures by which the age equivalences were obtained.

How are converted scores reported?

Age equivalent (Mental age), a standard score equivalent (I.Q.), and a percentile equivalent.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)

The ages of the norming population ranged from 2 years 6 months to 18 years. The procedures utilized in drawing the sample should result in a diversity of ability levels. No indication is given of the use of other instruments to measure the I.Q. of the preschool population.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?

The norms were obtained in 1958. The norming group included only white children and youth residing in and around Nashville, Tennessee.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?

The procedures are clearly explained and tables are provided.

Is special training required to interpret the test scores?

No specific information is provided on the interpretation of the PPVT. General interpretations which reflect the relation between verbal intelligence and performance in other areas might be made.

Can this test be directly, clearly, immediately interpreted?

No.

Does the test provide guidelines for making decisions from a score or score interpretation?

General guidelines are provided.

USEABILITY

Is this test administered to individuals or groups?

Individuals.

How much time does administration take?

15 to 20 minutes.

Does the manual contain a discussion of the purpose, use and limitation of the test?

Yes; and it also includes references to studies in which the PPVT has been used.

Does the manual contain clear administration and scoring directions?

Clear and specific.

Does the manual provide a description of test development, analysis procedures, and normative groups?

Detailed descriptions are provided.

Can the test be administered by the teaching staff without extensive practice?

Yes.

How much time is necessary to prepare for the test administration?

The examiner must be familiar with the pronunciation of all the words.

RETEST POTENTIAL

Are alternate forms of the test avilable?
Yes.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
According to Howard Lyman, the reviewer in Buros' Sixth Mental Measurements Yearbook, the norms of the two forms are biased because they are based on the whole sample rather than on the half that took a given form first.

SUMMARY

The PPVT is a widely used measure of a child's receptive vocabulary with inference often being of the child's verbal ability. It is easy to administer and score. Its usefulness is restricted by the limited geographic, ethnic, and racial composition of the population on which it was normed. Since the PPVT is often used with children from diverse ethnic and racial backgrounds, the pictures should be more representative of this diversity. In the CSE-ECRC Preschool/Kindergarten Test Evaluations, the PPVT received the highest rating possible ("good") on the measurement aspects of the test and indications of the need for improvement ("fair") on the other aspects of the test. Caution must be exercised in interpreting I.Q. scores at the lower range because small differences in chronological age can yield large I.Q. differences.

Title: Pre-school Attainment Record (PAR)
(Research Edition 1966-67)

Author: E. A. Doll

Publisher: American Guidance Service
Publisher's Building
Circle Pines, Minnesota 55014

Description:

The PAR is an expansion of the early age levels of the Vineland Social Maturity Scale. It is designed to measure the child's usual behavior in these categories: ambulation, manipulation, rapport, communication, responsibility, information, ideation, and creativity. It is administered as a structured interview with an adult closely associated with the child. There are 14 age items in each category (one item for each 6 month age span). The PAR may be used to determine the developmental level of children for whom verbal intelligence tests are not appropriate, for example, children with sensory impairments, speech or language disabilities or cultural differences. The age range is 6 months to 7 years. The cost of $2.50 for 25 blanks and $1.35 for a manual. Cost figures are slightly dated. (1966-67)

VALIDITY

What does this test measure?

Levels of development, particularly motor, social, and language are measured.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

Each category contains two items per year; several items seem to be out of place. For example, most children of 2-5 years do not "cooperate" as indicated on the scale. There is an overlap among the categories. Though, in the manual, the author discusses the problem of overlap in the categories, some consideration might be given to combining and redefining the categories.

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

The reviewer in Buros 7th suggests that this instrument appears to be a good basic inventory of the behavior of infants and children. However, several problems, including the lack of validity and reliability data hinder the users of this instrument.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

For the most part, the items seem suitable with the exception of several items in the "Responsibility" section which are excessively authoritarian: "Minds; obeys" at 2.0 and "conforms" at 6.0, for example.

Are instructions to the child about what he is to do and why clear and suitable in tone?

The child is not usually "tested." Information is obtained from an interview with adult or by an observation.

Are the printing and/or illustration of the appropriate style, size, and clarity? (clear, simple)

Yes.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

The fold-out record sheet might be more conveniently packaged as a booklet.

Are the length of the test and the time pressures appropriate?

The structured interview resulting from the use of the PAR will probably be of reasonable length.

How are answers recorded?

The answers are recorded by the examiner on the record form.

What equipment is needed?

Pencil.

Is it readily available?

Yes.

How much maintenance does it require?

None.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

The author states that "appreciable work" has been done on the PAR. He does not specify the work, however. Little indication of the method for placing items in the scale is given with the exception of a statement that items were placed according to information already available in the field.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)

None is indicated.

Are procedures for replicating statistical procedures clearly described?

The author describes procedures which should be utilized in the standardization of the PAR. He notes that much research is necessary on contingent variables. However, the author states that comparative studies can be made without normative standardization.

How are converted scores reported?

Raw scores are converted to "Attainment Ages" by dividing the raw score by .16, the total number of items per year. The scores may be displayed on a profile.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)

None is indicated.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?

None is indicated.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?

No indication is given for the number of items necessary to pass the category.

Is special training required to interpret the test scores?

Yes. The manual contains only a few statements about general interpretation.

Can this test be directly, clearly, immediately interpreted?

No.

Does the test provide guidelines for making decisions from a score or score interpretation?

Only general instructions are given.

USEABILITY

Is this test administered to individuals or groups?

Individuals.

How much time does administration take?

The PAR is untimed.

Does the manual contain a discussion of the purpose, use and limitation of the test?

All are discussed.

Does the manual contain clear administration and scoring directions?

The administration directions are clear. However, because the interview format is complex, more examples of "probe" questions are needed. The record form contains only brief description of the behavior. The examiner must obtain descriptions of the child's ordinary behavior attainments without asking leading questions.

Does the manual provide a description of test development, analysis procedures, and normative groups?

No.

Can the test be administered by the teaching staff without extensive practice?
The examiner should be trained in conducting an effective interview.

How much time is necessary to prepare for the test administration?
The examiner should become familiar with the definitions of the behaviors listed in each category.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

If a trained examiner can administer the PAR and interpret the information obtained about the child, then the PAR could possibly be used to determine the usual behavior of children for whom verbal intelligence tests may not be appropriate. The generally "poor" ratings received by the PAR in the CSE-ECRC Preschool/Kindergarten reflects both the lack of reliability and validity information and the difficulty of administration.

Title: Santa Clara Inventory of Developmental Tasks

Author: William Garner, Editor

Publisher: Richard L. Zweig, Associates
520 Richey Avenue
West Collingswood, New Jersey 08107

Description:

The Santa Clara Inventory of Developmental Tasks is designed to measure a child's development in motor coordination, visual-motor performance, visual perception, visual memory, auditory memory, and conceptual development. The teacher may observe many of the 60 tasks on the Profile during the regular teaching program. Each task is specified by the material, procedure, and scoring necessary to complete the task. The Inventory includes an Observation Guide which describes how to measure each task, a resource manual which contains activities for reteaching each task, 21 spirit masters on which the child performs certain tasks, and illustrated task cards. The upper limit of the Inventory is 7 years. The lower limit is not specifically indicated, but may be estimated as approximately 2 years from the items on the Inventory. The cost of the Inventory, including all materials, is $40.00. (1974)

VALIDITY

What does the test measure?

The Inventory measures motor, visual, auditory, language, and conceptual skills.

For the ages indicated, how well does the test measure what it purports to measure? (breadth and depth of items)

If used as a screening device, the Inventory can provide some useful information to the teacher about a child's development. The Inventory contains only 5 tasks in "Language Development" and 4 in "Conceptual Development."

What empirical evidence exists that would give an administrator confidence in its useage? (test utilization in experimentation, number and quality of content analyses, ability to discriminate between groups, evaluation in Buros' 7th or measurement journals)

In the preface, the publisher notes that the performance skills are considered by most authorities to be prerequisite to success in more advanced cognitive tasks. No specific references are listed nor is any other reference made to the validity of the Inventory.

EXAMINEE APPROPRIATENESS

Are the types of items suitable for the intended users?

Yes. The child would probably find most tasks interesting

Are instructions to the child about what he is to do and why clear and suitable in tone?

The directions to the child about what he is to do is clear. No indication is given of explanation given to the child concerning why he is performing a given task.

Are the printing and/or illustrations of the appropriate style, size, and clarity? (clear, simple)

Some of the illustrations, such as the one of the children in line are cluttered and somewhat inappropriate. Adult faces are attached to children's bodies. The clowns are too cluttered also.

Is the item packaging, use of white space, item-stem response coherence, and use of color effective and appropriate?

All the materials are packaged in a loose-leafed notebook. If the notebook were slightly larger, the pages would be easier to handle.

Are the length of the test and the time pressures appropriate?

The test is untimed; the teacher will not need to go through the entire Inventory at one time.

How are answered recorded?

The examiner shades in a block on the Profile which corresponds to the child's score on a task.

What equipment is needed?

Most materials, or spirit masters for producing the materials are included. A few materials, such as beads, must be assembled locally.

Is it readily available?

Yes.

How much maintenance does it require?

Minimal.

RELIABILITY, REPLICABILITY, REFINEMENT

What is the correlation between two or more administrations of the same test to the same or similar populations with suitable time intervals between testing periods? (test-retest reliability should be at least .70)

No measures of the reliability of this Inventory are reported.

What is the correlation between two comparable or alternate forms of the test? (equivalence reliability should be at least .70)

No alternate forms are available.

Are procedures for replicating statistical procedures clearly described?

No statistical procedures are indicated.

How are converted scores reported?

No converted scores are reported.

TEACHING FEEDBACK

What is the norm range? (more than three years in age, 3 years difference in education, and a variety of ability levels)

No information is presented.

Are the norms current (within 5 years) and representative of geographic areas, ages, racial and ethnic origins, types of schools and school districts?
No information is presented.

Are procedures for obtaining a raw score and converting the score (where appropriate) clearly explained and easy to execute?
The procedure for obtaining a raw score is clear but no information about the interpretation of the scores is provided.

Is special training required to interpret the test scores?
No indication is given about the interpretation of the scores.

Can this test be directly, clearly, immediately interpreted?
No. Remedial activities are provided for each area. However, the teacher would need skills in task analysis in order to select appropriate remedial activities.

Does the test provide guidelines for making decisions from a score or score interpretation?
No.

USEABILITY

Is this test administered to individuals or groups?
Individuals.

How much time does administration take?
Since the entire Inventory is not administered at one time, the amount of time is difficult to estimate and will vary with individual children.

Does the manual contain a discussion of the purpose, use and limitation of the test?
Only minimal information is provided, particularly about the limitations of the Inventory.

Does the manual contain clear administration and scoring directions?
The administration and scoring directions are exceptionally well presented. For each task, the material and procedures are listed for each task. Directions for scoring each item, along with examples, are included with each task rather than in a separate section.

Does the manual contain a description of test development, analysis procedures, and normative groups?
None is indicated except that the material has been used as a part of the Santa Clara Unified School District Title I project since 1967.

Can the test be administered by the teaching staff without extensive practice?
Yes. The teachers should be trained in observational skills, however.

How much time is necessary to prepare for the test administration?
Minimal.

RETEST POTENTIAL

Are alternate forms of the test available?
No.

Are alternate forms comparable in terms of content, validity, descriptive data, and reliabilities?
Not applicable.

SUMMARY

The strong points of this Inventory include the specification of the tasks, the provisions of examples to aid in scoring, the inclusion of most of the needed materials. Though several areas of visual and auditory skills are tapped, there are insufficient items in the language and conceptual areas. The lack of validity and reliability information makes this Inventory inappropriate for diagnostic evaluation. However, in the hands of a teacher with some training in early childhood education, this Inventory could yield information useful in making instructional decisions about children.

ADMINISTRATION

	Format	Number	Time	Training	Interpretation	Activities
ABC Inventory	A	I	8-9 min	No	No	No
ABR	A/0	I	untimed	No	Yes	No
Basic Concept Inventory	A	I	20 min	No	No	Yes
Carolina Developmental Profile	A	I	untimed	No	No	No
Circle Preschool Individual Assessment	A	I	untimed	No	No	Yes
CIP	A/I	I	30 min	Yes	No	No
CIRCUS	A	G	untimed	No	No	No
Comprehensive Developmental Scale (Project Memphis)	A/0	I	untimed	No	Yes	Yes
Cooperative Preschool Inventory	A	I	15 min	No	Yes	No
Denver	A	I	untimed	No	Yes	No
Developmental Progress Scale	A	I	untimed	No	Yes	No
Developmental Screening Questionnaire	0	I	untimed	No	No	No
DIAL	A	I	30 min	No	No	No
Koontz	0	I	one week	No	No	Yes
LAP	A/0	I	untimed	No	No	Yes
Peabody Picture Vocabulary Test	A	I	20 min	No	Yes	No
PAR	I	I	untimed	Yes	Yes	No
Santa Clara	A/0	I	untimed	No	No	Yes

Format: 0 = observation
I = interview
A = administer to child

Training: Is special training required to administer the test?

Interpretation: Is special training, such as a psychologist might receive, required to interpret the results?

Activities: Are activities provided for planning a program for the child?

Number: I = individual
G = group

	AGE LEVEL IN MONTHS						
	0-12	13-24	25-36	37-48	49-60	61-72	73-84
ABC Inventory				40	X	X	X
ABR		X	X	X	X	X	
Basic Concept Inventory					X	X	X
Carolina Developmental Profile			X	X	X		
Circle Preschool Individual Assessment		X	X	X	X	X	
CIP				30		66	
CIRCUS					X	X	
Comprehensive Developmental Scale (Project MEMPHIS)			X	X	X		
Cooperative Preschool Inventory				X	X	X	
Denver	2 wks	X	X	X	X	X	
Developmental Progress Scale	X	X	X	X	X	X	X
Developmental Screening Questionnaire				X	X	X	
DIAL			30	X	X	X	65
Koontz	X	X	X	X			
LAP	1	X	X	X	X	X	
Peabody Picture Vocabulary Test		30	X	X	X	X	X
PAR	6	X	X	X	X	X	X
Santa Clara		X	X	X	X	X	X

RELIABILITY, VALIDITY, NORMS

	Reliability	Validity	Norms
ABC Inventory	NI	NI	NI
ABR	NI	NI	NI
*Basic Concept Inventory	NI	NI	NI
*Carolina Developmental Profile	NI	NI	NI
Circle Preschool Individual Assessment	NI	NI	NI
*CIP	INSUF	RS	INSUF
*CIRCUS		NI	NI
Comprehensive Developmental Scale (Project MEMPHIS)	NI	NI	NI
Cooperative Preschool Inventory	INSUF	NI	RS
Denver	INSUF	INSUF	RS
Developmental Progress Scale	INSUF	INSUF	NI
Developmental Screening Questionnaire	NI	NI	NI
DIAL	INSUF	Yes/RS	RS
Koontz	INSUF	INSUF	NI
*LAP	NI	NI	NI
Peabody Picture Vocabulary Test	I	I	RS
PAR	INSUF	I	NI
Santa Clara	NI	NI	NI

*Authors of criterion-referenced tests often do not view information on validity, reliability and norms as relevant to their instruments. An * indicates these instruments are criterion-referenced.
NI: none indicated
INSUF: insufficient data or procedures
RS: restricted sample
I: indicated studies are reported

	USE AND CONTENT									
	Screening	Assessment	Motor	Language	Cognition	Social	Self-help	Emotional	"Readiness"	Other
ABC Inventory									X	
ABR	X		X	X	X			X		
Basic Concept Inventory	X			X					X	
Carolina Developmental Profile	X		X	X	X					
Circle Preschool Individual Assessment		X	X	X	X	X	X	X		
CIP	X		X	X	X	X	X			vision-hearing
CIRCUS		X		X	X				X	
Comprehensive Developmental Scale (Project MEMPHIS)		X	X	X	X	X	X	X		
Cooperative Preschool Inventory	X			X	X					
Denver	X		X	X		X	X			
Developmental Progress Scale	X		X	X		X	X			
Developmental Screening Questionnaire	X		X	X		X	X	X		
DIAL	X		X	X	X					
Koontz		X	X	X		X				
LAP		X	X	X	X	X	X			
Peabody Picture Vocabulary Test		X		X						
PAR		X	X	X	X	X	X			
Santa Clara		X	X	X	X					

REFERENCES

Anastasi, Anne. Psychological testing. New York: McMillan Publishing Co., Inc., 1976.

Bloom, B. S., Hastings, J. T. and Madaus, G. F. Handbook on formative and summative evaluation of student learning. New York: McGraw Hill, 1971.

Boyer, E. G., Simon, A. and Karafin, G. Measures of maturation: An anthology of early childhood instruments. Philadelphia: Humanizing Learning Program, 1973.

Cronback, L. J. and Gleser, G. The tester and decision theory. In J. T. Flynn and H. Garber, Assessing behavior: Readings in educational and psychological measurement. Reading, Mass.: Addison-Wesley, 1967.

Ebel, R. L. The social consequences of educational testing. School and Psychologist, 1964, 331-334.

Ebel, R. Improving the competence of teachers in educational measurement. In J. T. Flynn and H. Garber, Assessing behavior: Readings in educational and psychological measurement. Reading, Mass.: Addison-Wesley, 1967.

Evaluation: A bibliography of tests for use with the young child. University of North Carolina, TADS, 1973.

Evans, E. Measurement practices in early childhood education. In R. W. Calvin and E. Laffiro, Preschool education. New York: Springer, 1974.

Flapan, D., Gunn, G. and Newbauer, P. Sources of information for developmental assessment. Psychosocial Processes, 1970, 2, 98-115.

Haepfner, R., Hemenway, J. and DeMuth, J. et al. CSE-RBS Test evaluations: Tests of higher order cognitive, affective, and interpersonal skills. Los Angeles: Center for the Study of Evaluation, UCLA, 1972.

Hoephner, R., Stern, C. and Nummedal, S. CSE-ECRC Preschool/kindergarten test evaluations. Los Angeles: Center for the Study of Evaluation, UCLA, 1972.

Joint Committee for Standards for Educational and Psychological Tests and Manuals. Validity and reliability defined. In J. T. Flynn and H. Garber, Assessing behavior: Readings in educational and psychological measurement. Reading, Mass.: Addison-Wesley, 1967.

Horrocks, J. E. Assessment of behavior. Columbus, Ohio: Charles E. Merrill, 1964.

Kamii, C. E. Preschool education: socio-emotional, perceptual-motor, cognitive development. In Bloom, et al., Handbook on formative and summative evaluation of student learning. New York: McGraw Hill, 1971.

Meier, J. Screening and assessment of young children at developmental risk. Washington: DHEW #(05)73-90, 1973.

Medinnus, G. R. Child Study and Observation Guide. New York: John Wiley and Sons, Inc., 1976.

Olin, T. and Rodabaugh, J. A selected listing of instruments for screening and assessment: Bibliography for children with handicaps. Office of Child Development of Region V, Chicago.

Rowen, B. The children we see; an observational approach to child study. New York: Holt, Rinehart, and Winston, Inc., 1973.

Snelbecker, G. E. Learning theory, educational theory, and psycho-educational design. New York: McGraw Hill, 1974.

Tyler, L. Tests and measurements. Englewood Cliffs, N. J.: Prentice Hall, 1971.

Thorndike, R. L. Decision-making and prediction. In J. T. Flynn and H. Garber, Assessing behavior: Readings in educational and psychological measurement. Reading, Mass.: Addison-Wesley, 1967.

Thorndike, R. L. Educational measurement (2nd ed.). Washington: American Council on Education, 1971.